Remarks/Arguments:

Claims 1-15 are pending.

Applicants acknowledge with thanks the courtesy shown to their representative by Examiner Armstrong during the telephone interview of March 3, 2008. During the course of the interview, Applicants' representative explained the features of Applicants' claim 1 with respect to Figs. 2 and 8A of the subject specification. In addition, Applicants' representative discussed the differences between the cited art and Applicants' claim 1. No agreement was reached.

Applicants' invention relates to methods for speaker normalization and apparatus for speech recognition. Applicants' claims include features neither disclosed nor suggested by the cited art. Namely, the cited art do not disclose or suggest the combination of: 1) determining, for each frame, a plurality of similarities/distances using frequency-converted feature parameters, 2) selecting at least one frequency conversion coefficient using the plurality of similarities/distances for each of the frames and 3) normalizing the input utterance by frequency-converting the input utterance using the selected frequency conversion coefficient. First, an explanation of these claimed features is provided with respect to Fig. 8A of the subject specification. Following the explanation of the claimed features, differences between Applicants' claimed features and the cited art are described.

Claims 1-15 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over Yamada et al. (U.S. 5,692,097) in view of Chuang (U.S. 4,941,178). This ground for rejection is respectfully traversed for the reasons set forth below.

Claim 1 includes features neither disclosed nor suggested by the cited art, namely:

> ...for each of the frames, frequency-converting the respective acoustic feature parameter by filtering with a plurality of predetermined frequency conversion coefficients to form a corresponding plurality of frequency-converted feature parameters...
>
> ...determining, for each frame, a plurality of similarities or distances between each of the frequency-converted feature parameters and a standard phonemic model...
>
> ...selecting at least one of the plurality of predetermined frequency conversion coefficients... by using the determined plurality of similarities or distances for each of the frames...

> normalizing the input utterance by frequency-converting the input utterance using the selected at least one predetermined frequency conversion coefficient... (Emphasis Added)

Claim 8 includes a similar recitation.

The features of claim 1 and 8 are next described with reference page 17, line 2-page 18, line 25 and Figs. 8A, 8B and 9B of the subject specification. For each frame, a respective acoustic feature parameter is filtered with a plurality of frequency conversion coefficients, for example, $\alpha_1,\ldots, \alpha_7$. In Fig. 8A, in each frame, a maximum likelihood 801 (i.e. similarity/distance) of a conversion coefficient α_i is selected for each phoneme within the frame. For example, in the first frame, α_4 is selected with /a/, α_3 with /e/, etc. Thus, for each frame, a plurality of similarities/distances are determined between each of the frequency converted feature parameters and a standard phonemic model. Next, for each frame, one maximum likelihood 802 of a phoneme and a corresponding conversion coefficient 803 are determined from among all similarities/distances 801. For example, for the first frame, the one maximum likelihood 802 is /a/ and the corresponding conversion coefficient is α_4.

At least one frequency conversion coefficient is selected, using the determined pluralities of similarities/distances (801) for each frame. For example, as shown in Fig. 8B, conversion coefficients 803 are compared over all frames and a most frequently occurring conversion coefficient is selected (for example, α_4). As another example different conversion coefficients may be selected for each frame (Fig. 9B). The selected conversion coefficient(s) are used to normalize the input utterance. Thus, claims 1 and 8: 1) filter the acoustic feature parameter, for each frame, with a plurality of frequency conversion coefficients α_i, 2) determine a plurality of similarities/distances for each frame (801) and 3) select at least one of the frequency conversion coefficients based on the similarities/distances (801). The input utterance is normalized by frequency-converting, using the selected frequency conversion coefficient.

Yamada et al. disclose, in Figs. 1 and 6, a voice recognizing apparatus including feature parameter extracting unit 13 that extracts feature parameters, phoneme similarity calculating unit 15 and normalized similarity vector calculating unit 16. Phoneme similarity calculating unit 15 determines a phoneme similarity for each frame between standard pattern phonemes (in storing unit 14) and the extracted feature parameters (from feature parameter extracting unit 13) to obtain similarity vectors (Col. 2, lines 4-10 and Col. 2, line 61-Col. 3, line 26).

Normalized similarity vector calculating unit 16 normalizes a vector length of each similarity vector to unity (Col. 3, lines 46-50).

As acknowledged by the Examiner on page 2, paragraph 4 of the Office Action, Yamada et al. do not disclose or suggest Applicants' claimed feature of "frequency-converting the respective acoustic feature parameter by filtering with a plurality of predetermined frequency conversion coefficients to form...frequency-converted feature parameters" (emphasis added). In addition, Yamada et al. do not disclose or suggest Applicants' claimed features of 1) "determining...a plurality of similarities or distances between each of the frequency-converted feature parameters and a standard phonemic model," 2) "selecting at least one of the...predetermined frequency conversion coefficients...by using the determined plurality of similarities or distances for each of the frames" or 3) "normalizing the input utterance by frequency-converting the input utterance using the selected...predetermined frequency conversion coefficient" (emphasis added). These features are neither disclosed nor suggested by Yamada et al.

On page 2, paragraph 4 of the Office Action, the Examiner asserts that Yamada et al. discloses a voice recognition method for recognizing a word in speech, which implements a normalizing similarity vector calculating unit, and refers to Col. 18, line 18-Col. 31, line 44 of Yamada et al. Applicants have carefully reviewed Col. 18, line-Col. 31 line 44 of Yamada et al. and can find no disclosure of Applicants' claimed features of: 1) determining a plurality of similarities or distances between each of frequency-converted feature parameters and a standard phonemic model, 2) selecting at least one predetermined frequency conversion coefficient using the determined similarities or distances for each of the frames and 3) normalizing the input utterance by frequency-converting the input utterance using the selected predetermined frequency conversion coefficient. Accordingly, Applicants respectfully request that the Examiner either specifically point out where Yamada et al. disclose these features or withdraw the rejection.

Chuang discloses, in Fig. 1A, a speech recognition system including slope filter estimate 16 and inverse filter 22 that provide a slope removal process to normalize the slope of LPC coefficients (Col. 4, lines 1-26 and Col. 6, line 63-Col. 7, line 52). The speech recognition system also includes all-pass filter 30, spectral warping 32 and time warping 34 for spectral normalization (after the slope normalization) and where the slope normalization and spectral normalization are regarded as speaker normalization. (Col. 8, lines 15-61 and Col. 9, lines 60-

62). All-pass filter 30 provides expansion and compression of the LPC analysis results 24 along the frequency axis (Col. 8, lines 15-31 and Col. 8, lines 62-Col. 9, line 37).

Chuang does not make up for the deficiencies of Yamada et al. because it does not disclose or suggest: 1) determining a plurality of similarities or distances between each of frequency-converted feature parameters and a standard phonemic model, 2) selecting at least one predetermined frequency conversion coefficient by using the determined similarities or distances for each of the frames or 3) normalizing the input utterance by frequency-converting by the input utterance using the selected predetermined frequency conversion coefficient, as required by claim 1. Applicants have reviewed Col. 8, line 15-Col. 9, line 37 of Chuang, cited by the Examiner, and can find no disclosure of these features of claim 1. Applicants respectfully request that the Examiner either specifically point out where Chuang discloses these features or withdraw the rejection. As described above, the combination of: 1) determining, for each frame, a plurality of similarities/distances between frequency-converted feature parameters and a standard phonemic model, 2) selecting at least one predetermined frequency conversion coefficient using the determined similarities/distances for each of the frames and 3) normalizing the input utterance by frequency-converting the input utterance using the selected predetermined frequency conversion coefficient is neither disclosed nor suggested by the cited art. Accordingly, allowance of claim 1 is respectfully requested.

Claims 2-7 include all of the features of claim 1 from which they depend. Accordingly, claims 2-7 are also patentable over the cited art.

Claim 8, although not identical to claim 1, includes features similar to claim 1 that are neither disclosed nor suggested by the cited art. Namely, 1) frequency converting an extracted acoustic feature parameter by filtering with a plurality of predetermined frequency conversion coefficients, 2) determining plural similarities/distances with frequency-converted feature parameters, 3) selecting at least one predetermined frequency conversion coefficient by using the determined similarities or distances for each frame or 4) normalizing the input utterance by frequency-conversion using a selected frequency conversion coefficient. As discussed above, these features are neither disclosed nor suggested by the cited art. Accordingly, allowance of claim 8 is respectfully requested.

Claims 9-15 include all of the features of claim 8 from which they depend. Accordingly, claims 9-15 are also patentable over the cited art.

In view of the arguments set forth above, the above-identified application is in condition for allowance which action is respectfully requested.

Respectfully submitted,

Lawrence E. Ashery, Reg. No. 34,515
Attorney for Applicant

DMG/nm

Dated: March 25, 2008

P.O. Box 980
Valley Forge, PA 19482
(610) 407-0700

NM237783